

09/831139

GR 98 P 4724

1/PRTS

JC08 Rec'd PCT/PTO 07 MAY 2001

XNS A1
Description

5 Data processing system or communications terminal with a device for recognizing speech and method for recognizing certain acoustic objects

Devices and methods for recognizing natural speech are today familiar to a person skilled in the art from many different applications. The practical 10 applicability and capacity of systems of this type depends very much on their complexity and the extent of their range of applications. The general principle applies that the recognition rate of such a system usually decreases greatly with an increasing number of 15 acoustic objects to be recognized (words, phonemes, individual letters, etc.). At the same time, however, measured in terms of cost and space requirement but also with regard to training effort, the expenditure 20 also usually increases greatly with the extent of applications.

Conventional speech recognition systems are therefore still not used for many applications, although in principle they would be suitable for them from the viewpoint of the user. The invention is therefore based on the object of specifying a technical teaching which makes it possible for speech recognition to be used even for those applications where relatively great expenditure has to be ruled out for economic or other reasons. This object is achieved by a data 30 processing system or communications terminal with a device for recognizing speech or by a method for recognizing certain acoustic objects according to one of the patent claims.

The product according to the invention, a data processing system or communications terminal, has a device for recognizing speech which is set up specifically to recognize certain acoustic objects, to be specific individual letters, combinations of letters or control commands, or can be specifically configured

GR 98 P 4724

- 1a -

to recognize such objects.

卷之三

The same applies correspondingly to the speech recognition algorithm of a method according to the invention. Furthermore, a device for the acoustic output or optical display of recognized acoustic objects is provided. In this way, the number or set of the acoustic objects to be recognized can be largely adapted to the intended application. The envisaged device for the acoustic output or optical display of recognized acoustic objects makes possible a direct feedback between the user and the system, providing the user with effective control over the recognition capacity and allowing the number of misrecognitions to be reduced in a simple but very effective way.

If the user establishes a misrecognition on the basis of the acoustic output or optical display, he can repeat the acoustic input of the object to be recognized. Since this process possibly does not lead to correct recognition in a very short time, it is provided according to a preferred embodiment of the present invention that the speech recognition device is set up or can be configured in such a way that the recognition of a certain first control command has the effect following the output or display of an acoustic object of triggering the output or display of a further acoustic object. This enables the user after the output or display of an acoustic object, that is for example after an established misrecognition, to make the system output a further acoustic object by the acoustic input of a special acoustic object, to be specific a control command.

If, for example for a selection $\{AO_1, AO_2, \dots, AO_n\}$ of possible acoustic objects, the device for speech recognition or the speech recognition algorithm determines recognition probabilities $\{p_1, p_2, \dots, p_n\}$ with the property $1 > p_1 \geq p_2 \geq \dots \geq p_n > 0$, this preferred

embodiment makes possible, for example, the output or display of AO2 after the output of the misrecognized object AO1, or similar measures for supporting a correction of the recognition error that is as convenient as possible for the user. A possible selection for such a special acoustic object or such a control command would be, for example, the word "incorrect". It is not difficult for a person skilled in the art to consider on the basis of this description further application possibilities for this embodiment of the present invention.

Further preferred embodiments of the present invention are the subject of further subclaims.

The invention is explained in more detail below on the basis of preferred exemplary embodiments with the aid of figures.

Figure 1 shows in a schematic way the structure and mode of operation of a preferred embodiment of a system according to the invention.

As represented in figure 1, this embodiment of a data processing system (DPCD) or communications terminal (DPCD) according to the invention comprises a speech recognition unit (SRU), which recognizes acoustic objects (AO) spoken by a user of the system and feeds the recognized acoustic objects (RAO) to a device for acoustic output or optical display (DU). According to the present invention, the speech recognition device is set up specifically to recognize certain acoustic objects (AO), to be specific individual letters, combinations of letters or control commands, or can be configured specifically to recognize such objects.

The speech recognition device consequently assigns to an acoustic object (AO) spoken by the user in each case an acoustic object recognized by this device (RAO). Since the recognition of natural speech is always

subject to a certain uncertainty for fundamental reasons, the recognized acoustic object will generally be, depending on the speech recognition algorithm used, the most probable or most plausible acoustic object
5 that comes into consideration, taking into account the determined features of the spoken acoustic object.

The user receives via the output or display device (DU) an acknowledgement message concerning the result of the recognition process. He then has the possibility of responding to this according to the type of result involved. If the acoustic object was misrecognized, he has the possibility of notifying the speech recognition algorithm that the acoustic object has not been correctly recognized, or that he wanted to
15 have a different object recognized, by saying a control command intended for this purpose, for example the word "again". He then has the opportunity to say once again the object desired by him. This process can be continued until the speech recognition unit recognizes
20 the desired object.

The input of another control command, for example the word "incorrect", could control the speech recognition algorithm in such a way that a further acoustic object is output, preferably that object of which the probability or plausibility is admittedly lower than that of the object previously output but greater than that of all the other objects coming into consideration. In this case, it would not be necessary for the user to say the object again; instead, further
30 candidates would continue to be offered for the object to be recognized until the user no longer inputs the corresponding control command or possibly inputs an expressly confirmatory command, for example "correct".

According to a further preferred embodiment, it
35 is possible to provide a control command, for example the word "continue", which, when recognized following the speaking or display of an acoustic object, has the effect of triggering the display or output of an

object which follows the former object in a certain sense. The sequence of the objects does not in this case have to be fixed by the magnitude of recognition probabilities or plausibility values but may also be
5 dictated by the sequence of entries in a memory unit (MU) of the system, or by alphabetical sequences of objects or sequences of objects semantically defined within a defined context. For example, the sequence of objects could be defined by the order within a
10 database, a telephone directory or the structure of a file stored in the memory unit, for example a customer file, a dictionary, or similar files.

JNS
15 XTD

When this patent application mentions devices which are set up or can be configured for a certain function or mode of operation, this means that the corresponding functional features of these devices may be permanently or temporarily restricted. Furthermore, these devices can be set up or configured by all those involved between the manufacturer and the user by
20 manufacturing processes, settings on the hardware or the use or parameterization of software or equivalent means or measures for a certain function or mode of operation. A person skilled in the art will readily deduce from this description numerous similar or
25 equivalent means or measures for this purpose.

A speech recognition device is preferably set up or configured by a suitable selection or parameterization of the software which realizes the desired function in the speech recognition algorithm
30 and/or the sequence control of this device. A data memory is preferably set up or configured by a suitable selection or parameterization of the data structure, for example the database structure, which defines the type of storage of the data on this memory and the type
35 of access to these data.

The effective recognition capacity of the system can be distinctly improved by the recognition of an acoustic object or a sequence of objects which corresponds or correspond to an entry in the data 5 memory having the effect of triggering the display or output of this entry (ME) or a function (FU) of the system associated with this entry. As a result, the existing prior knowledge of the objects likely to be recognized can be utilized very advantageously. 10 Although this technique is known in principle to a person skilled in the art, it is particularly effective, as appropriate tests have shown, in connection with a speech recognition system specially designed to recognize a limited set of objects to be 15 recognized, for example individual letters.

So if, for example, the first three letters of an entry in a telephone directory are recognized, a preferred embodiment of the invention provides the output or display of this telephone directory entry. 20 If it is not the desired entry, it may be sufficient to input (i.e. say) a control command or a few further control commands, such as for example "continue" or "street" or "fax number" or "connect", to achieve on the basis of, for example, the name of a subscriber 25 known to the user the output of the latter's fax number or the dialing of this number by the communications terminal by saying the first three initial letters of his name. Other functions which could be triggered in this way, such as for example the output of a text or 30 image, the display of a data record, etc., are so numerous that it is not possible to list them here.

The capacity of the systems or methods which realize the present invention can be further increased by providing certain control commands, such as for 35 example "letter", "control" or "combination", etc., the speaking of which enables the user to restrict the set of

objects to be recognized according to his choice (temporarily or permanently) to a certain subset, such as for example individual letters, combinations of letters or control commands.

With the present invention, in particular the number of telephone entries which can be called up by voice selection in a mobile telephone or cordless phone or in a wire-bound telephone can be increased at will. In the case of customary systems of this type, only a limited number of entries was allowed for voice selection, from experience at most 20 or 30 entries. This was due to the memory space to be made available for the voice samples to be re-recognized, i.e. due to the resultant costs and space requirement. If the number of entries was further increased, experience showed that the effort for training the speech recognition increased considerably, which led to lower user acceptance.

According to a preferred embodiment of the present invention, the speech recognition algorithm is trained by the user only for the letters of the alphabet, and possibly combinations, and just a few control commands. It is in this way set up or appropriately configured by the user for the recognition of these acoustic objects. Interrogation takes place by the acoustic input of initial letters and (preferably up to two) subsequent letters. Misrecognitions are reduced by plausibility checks, i.e. for example by comparison of the objects with entries in a memory device. The names input are spoken only once and converted in an encoder with a low bit rate (for example half rate of GSM) and stored at the corresponding memory location, possibly in a compressed form.

Alternatively, a synthesis program which synthesizes voice from a name may also be used, possibly requiring less memory space. In any event, the speech

recognition does not have to be trained for a large number of names but only for a fixed set of approximately 30 sequences of letters and control commands.

5 To use this embodiment of the invention, the user activates the service feature "voice selection", for example by means of the scroll key at the side, and inputs the first letters of the entry sought, possibly in the form "letter A" etc. Experience shows that the 10 probability of recognition is considerably greater in this case than in the case of a single letter. Each input is acoustically acknowledged by the recognized object being output. If the object was correctly recognized, the next object to be recognized is input.

15 If an object is recognized wrongly, the user responds with "incorrect" or "no". The system then proposes the next probable letter, for example instead of "D" a "T" or instead of "H" an "A" and so on. In most cases, it is sufficient to input the first two or 20 three letters to find the correct entry. If a corresponding control command is input or no further input takes place (control command = pause in speech), the terminal outputs the corresponding name in the telephone directory of the terminal. If there are a 25 number of entries with the same initial sequence of letters, the user issues, for example, the command "continue", until the "correct" name is acknowledged.

If a letter is recognized wrongly and, as a consequence, a first letter that is remote in the 30 alphabet - for example "T" instead of "D" - is output as the beginning of the input combination of letters, the user inputs (i.e. speaks) the control command "selection". The terminal then proposes the most probable next correct combination of initial letters. 35 Knowledge of the names stored in the telephone directory allows most possible wrong combinations to be ruled out from the outset. After that, the user issues the command "dial".